

国立国語研究所学術情報リポジトリ

言語処理におけるターンアラウンド・システム

著者	斎藤 秀紀
雑誌名	電子計算機による国語研究
巻	8
ページ	63-112
発行年	1977-02
シリーズ	国立国語研究所報告 ; 59
URL	http://doi.org/10.15084/00001046

言語処理における

ターンアラウンド・システム

斎藤 秀紀

0. 序

国立国語研究所（以下国研）では、各種の用語用字調査を進めてきたが、調査規模の拡大と期間の短縮を図るため、昭和41年に HITAC-3010 (20KC) 型コンピュータを導入した。これらのコンピュータ・システムは、同年に発足した、「新聞の語彙調査」に使用され、以後、並行して言語の自動処理研究、用例作成等に利用されてきた。しかし、これらのシステムは、磁気テープを主体とした装置であり、記憶容量も比較的小さかったため、昭和49年に「高等学校教科書の用語調査」として発足する調査には、新型コンピュータ HITAC-8250 (98KB) に機種変更した。また、翌50年には、漢字テレタイプ附属モニタ印字装置に代わるものとして、高速漢字プリンタ NEACC5210D を導入した。

本稿では、これらのシステムの導入に関して、検討の対象となった、基本的事項、システム機能、漢字プリンタと光学記号読み取り装置によるターンアラウンド処理、漢字ディスプレイによるデータ校正処理の概要を述べ、これらのシステムが用語調査に対し、効果的に適用できることを示すとともに、従来のシステムとの相違点と適用可能な応用面について概説する。

1. 漢字プリンタ導入の背景

昭和41年に HITAC-3010 と共に導入した、漢字入出力装置は印字速度は毎分 120 字と非常に遅く、収容字種も総数2400字に制限された装置であった。漢字プリンタの導入は、これらの装置の低速を補うことと印字字種の拡張を目的

に計画されたものである。しかし、システムの検討を進めるうちに、汎用の漢字処理システムと、より大きな機能を持ったシステムへと要求が拡大していった。一つには、従来のコンピュータ処理方式が大量処理向きのバッチ処理形態であったため、かならずしも、研究者の研究上の要請と一致していなかったことによる。ターンアラウンド処理機能は、このような背景のもとで計画され、各装置は可能な限り、この線にそって構成されている。しかし、これら漢字プリンタ、OMR 装置によるターンアラウンド処理については、装置の開発の方向づけが明確になされていないため、各装置間のプリンタによるマーク位置と OMR 装置の記号検出位置の調整、静電記録用紙による OMR シートの作成、またシートが作業帳表として十分使用可能であることが大きな問題となった。以上の点について、我々が要求した基本機能は次の五項目であるが以下個々の機能と問題点について述べる。

- 1) 高速印字機能
- 2) 任意のコード体系への変換機能
- 3) 漢字ディスプレイによる情報の検索と校正機能
- 4) 外字挿入の容易性
- 5) OMR装置及び漢字プリンタ間でのターンアラウンド機能

2. 漢字プリンタと制御用コンピュータ

漢字プリンタシステムは、内蔵ミニコンピュータ N-3200 によって制御されるスタンドアローン方式の制御形式をとる(機器構成は図 3 参照)。通常、プリンタ等の単体制御用コンピュータは「装置の制御」に、その目的が限定されているため、小型機が実装されている場合が多い。これは、装置の使用目的が単能機として使用目的がしばられているため、コンピュータはその背後にかくれ、いわゆるコンピュータそのものとしての性格は薄れる。その結果、システムは機能印字にのみ重点がおかれ、このために必要な編集、コード変換等の前処理は全てホストコンピュータによって処理することを前提に構成されること

が多くなった。また、メーカー側もそのような考え方を基本としてシステムを構成していたことは否めない。しかし、印字処理に欠くことのできない、外字処理または漢字登録テーブル等のメンテナンスについても、メーカーのセンタマシンを使用するか、新たにホストコンピュータ用にプログラムを開発する必要が生じ、かならずしも運用面で満足できる状態ではない。これは、前述のようにプリンタの基本的機能が印字処理を指向したものであり、印字処理のためのバックアップ体制までプリンタ側でカバーすることを考えていないことが大きな要因となっている。

ミニコンピュータの性能については、従来、我国のコンピュータメーカーの大部分が汎用コンピュータの開発を主体としており、ミニコンピュータ専用メーカーの目ざす汎用性とは、異った思想のもとで開発されている。これは、あくまで、汎用コンピュータの下位機種として上位機種と競合しないよう、使用範囲が制限されている場合が多い。内蔵ミニコンピュータの手足の弱さ、記憶容量の制限は、必然的にシステムに必要な装置の接続能力、開発するプログラムの拡張の可能性を自ら閉じてしまうことになる。漢字プリンタにおける処理について、従来の印字機能から使用者のニーズにあったシステム構成を考える場合、コンピュータ周辺装置の柔軟な使用が必要となる。また、コンピュータの周辺装置に対する接続能力とその制御能力に余裕があることは、将来の機器増設についても道を開くとともに柔軟な使用が可能となる。

本システムにおける OMR、漢字ディスプレイの接続、システムに対する基本的要求の五項目も、これらコンピュータの能力に負う部分が多い。漢字プリンタシステムに基本的処理の大部分を負わせることは、ホストコンピュータとの導入時点が異なる場合、またプログラム開発要員、ホストコンピュータに余裕がない場合等、ジョブの増加分に対し分散処理が可能となり、これらの機能はシステム運用上非常に重要であると思われる。

3. 漢字プリンタ印字方式と用紙

システムについて、ターンアラウンド処理の技術的問題は、汎用コンピュータ処理のアプリケーションとして一応解決されているにもかかわらず、漢字処理をそのまま適用するためには、問題が多い。特に、漢字処理の場合、そのアプリケーションとしての適用領域の開拓が確立されていないため、過去にターンアラウンドの有効性が実証されているにもかかわらず、装置の機能にあまり注意深い配慮がなされていない場合がある。この場合、メーカーの最初の出発点が大きく影響しているように思われるが、写植からの場合とコンピュータ用端末装置としての場合に微妙な違いがある。電算写植の場合、明らかに活字に変わるものとして、出版のための高文字品種を目ざしており、出力に力点がおかれている。一方、コンピュータ用端末の一種として開発された場合、多少の文字品質その他、機能を犠牲にしても、高速印字に重点がおかれていたように思われる。これらは、印字可能な文字のポイント、字間、行間等は印刷のレイアウト程度の機能にとどめられていた。しかし、コンピュータの入出力装置は先行した英数字処理に対して、全ての規格が定められており、各装置もまた、このために調整されている。ここで、漢字処理の中にターンアラウンド機能を取り入れる場合、まったく、別のものとして漢字に対する規格を定めるか、また、従来の入力装置の規格を利用可能にする、出力調整機能が要求される。

また、用紙については、現在のプリンタの開発が、おおむね普通紙印字方式方向に進められているが、まだいくつかの方式が使用されている。

表 1 印字方式と使用用紙

電子写真方式	ゼログラフィー方式	乾式	普通紙	32×32
	CANON NP 方式	乾・湿式	普通紙	32×32
	エレクトロファクス方式	湿式	酸化亜鉛紙	32×32
静電記録方式	多針電極記録方式	湿式	静電記録紙	24×24
インクジェット方式	インク		普通紙	32×32
インクミスト方式	インク・ミスト方式		普通紙	17×17
機械式 (インパクト方式)	ワイヤ、ヘリカルドット	リボン	普通紙	17×18

その代表的なものは表1のようなものである。

プリンタは高速で漢字を多量に印字する装置であるから、まず、用紙が安いこと、用紙の厚さを自由に選択できること、次に、印字文字がきれいなこと、装置として動作が安定しており、かつ装置価格が安いことが必要条件となる。用紙の価格の点では普通紙が最も安く、ついで、静電記録紙となっている。また、印字された文字が比較的解像度を高く維持できるのが電子写真方式であるが、ターンアラウンド処理を考える場合、印字後の紙の変形、汚れ、印字位置の正確な位置付け機能、紙送りの安定性が要求される。その他、プリンタ用に使用されている紙送り機構と用紙の形状については、装置の開発または導入時点で見おとされやすい点であるが、これはプリンタで出力された結果に対し、用紙ともに他の装置で再加工する場合、受け入れ側の装置の用紙の処理方式に当然のことであるが影響を受けることになる。一般に使用されているロール紙の場合、ロール状で出力される場合と用紙裁断用カットマークに従ってシート状で出力される場合がある。ロール状の場合、再加工の対象となる装置上にカットマーク検出装置が必要となり、これを任意の紙幅について要求されることになる。（プリンタが電子写真方式の場合は紙幅は9インチA4サイズに規定される場合が多い）また、シート状の場合は、裁断精度、受け入れ装置側の用紙の位置付けが問題となろう。同様に最初からシート状用紙を使用する場合も、用紙のプリンタ装置上でのポジショニングの問題が加わることになる。以上の点から、紙送りについては、従来からラインプリンタが使用されてきた、スプロケットによる紙送りが他の方式に比べて十分効果的であると思われる。また、ハード上で用紙のポジショニングに微調整機構を持っているため、用紙に対するポジショニングもより正確なものとなる。このような周辺的な問題についても、印字機構の開発も重要なものであることは言うまでもないが、漢字処理におけるターンアラウンド処理で問題となった入出力装置上の規格不統一と同様、プリンタの使用目的を広げる意味で配慮すべきポイントのひとつであると思われる。以上の点を考慮し、主に静電記録方式について、説明す

る。

多針電極による静電記録方式の場合、塩化ビニール系の誘電体を紙面に塗付したものを用紙として使用し、ドットに分解した文字イメージを電極上に電荷として与え、現像、定着を行うものである。文字の定着のさい、液剤使用と、紙面の片面に誘電体を塗付するため紙のカールと液剤のにじみが問題となる。また、紙が厚くなるに従って、トナーのカブリが増えることも問題の一つとなる。しかし、静電記録方式の場合、他の印字方式、例えばPPC方式に比べ、印字機構は物理的な電極そのものであり、簡単で安定度も高く、比較的高解像度を維持でき、かつ高速印字が可能である。また、文字の大きさに関係なく、行あたりの印字文字数は電極の長さによって、調整できる。この電極による、文字の表現は、 24×24 のドット表示であるが、PPCの場合の解像度が10本/mmであるのに対し、12ポイント用電極6本/mm、9ポイント用電極8本/mmと、若干おとる。しかし使用する字体が、大部分の印字装置が明朝体の使用しているのに対し、特別にデザインされた、例えばタイポス体の使用によって十分カバーが可能である。これらは、将来、マイクロフィルム等への転写を行う場合、縮小、複写による解像度の低下に対しては、従来の字体を使用する場合より有利であろう。静電記録用紙については、新聞社等で使用されているテレックス用紙と同一であり、価格、供給ともに安定しており、紙質も55kg~135kgと普通紙なみのものが使用でき、普通紙と比較しても印字上の遜色は見られない。また、用紙送りが、スプロケット (NEAC-C 5210Dの場合) であるため、印字位置は正確に調整でき、ターンアラウンドのためのIDコード等の印字が可能である。また、紙厚は135kgの連続用紙が使用できることは、紙のカールが比較的少なく、OMRシート、80欄カードによるターンアラウンド処理も可能にしている。紙の変形は、ロール紙の場合もこの紙のカールが大きな問題となる。また、PPC方式については、熱定着のさいの用紙の変形、潜像保持媒体の劣化による用紙の汚れ、その他、潜像ドラムを使用するさい、ドラムの直径の小さいものを使用する場合、紙のカールに注意しなければならない。これによ

って、使用できる用紙厚の制限を受けることになる。

4. 漢字辞書とコード変換処理

漢字プリンタを使用し、印字を行うためには、基本的には、標準 NEC コードでデータが記録されていることが必要である。これに対し、現有の入力装置である漢字テレタイプはコード体系が旧型コンピュータ HITAC-3010用に設計されているため、コード変換処理が必要となる。変換処理の方法は、コンピュータによるソフトの処理、コード変換専用機、入出力装置にハード変換機能による方法等が考えられる。大部分の処理は、コンピュータによる変換が行われているが、一部、変換専用機、また、任意のコード体系で出力可能な漢字入力装置も開発されている。本稿では、漢字プリンタを汎用コンピュータ的に使用することによって、コード変換処理を行う方法、また、漢字文字発生装置の特種な使い方による半ハードの変換について述べ、あわせて、印字処理にいたる過程と漢字パターン辞書の構成内容について述べる。コード変換処理の必要性については、漢字コードの標準化が進められているにもかかわらず、各社独自のコードが使用されているのが現状であり、これらは、外部へのパンチさん孔委託のさいコード変換処理は必須のものとなる。

また、過去に蓄積されてきたデータコード、標準化の対象外となっている、いわゆる盤外字コードに対する標準化の問題等、いまだ未解決の部分が多い。以上の点から、任意のコード体系への変換機能を保持しておくことは、実際的な仕事を行う上でも不可欠であり、また、外部で作成されたデータや各調査機関で作成された資料への接続も開いたものとなる。今後とも限定された予算内で、調査とデータの総合化を図る上で、これらの機能はより有効なものとなろう。

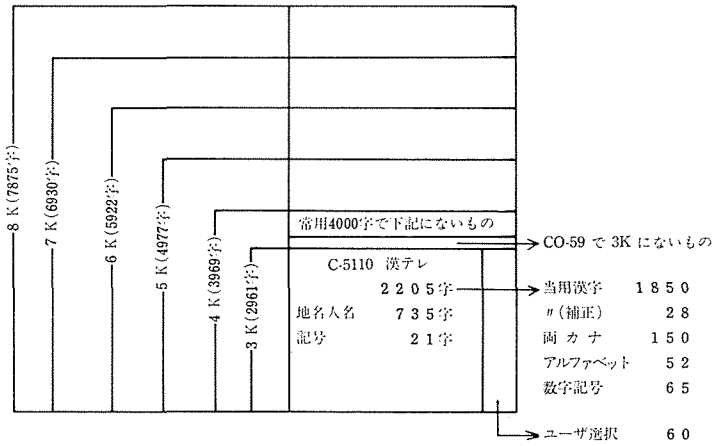
次に、我々の導入した、NEAC-C5210D 型漢字プリンタの印字処理について、漢字辞書（テーブルⅡ）の漢字文字発生装置（Read Only Memory：ROM）への登録の方法、また、コード変換テーブル（テーブルⅠ）との関係

について、簡単に説明する。印字処理の対象となる漢字は約 8 K 字であるが、比較的使用頻度の高いと思われる文字（図 1 キャラクタセット参照）5 K 字を高速度の ROM へ記録し、残りの 3 K 字を附属周辺の磁気ドラム上に格納する。漢字辞書におさめられている 8 K 字の内容は、NEC から提供されているものを基本としているが、構成は、次の通りである。使用頻度の高い、当用漢字補正漢字を基本とし、これらを 3 K 字以内におさめる。以下、使用目的によって異なるが、現在標準セットとして発表されている文字セットの中から、共同通信社の CO-59、活字工業会常用漢字 4000 字、情報処理学会漢字コード委員会試案 6100 字を参考として、7 K 字までの領域にこれらの字種が全て含まれるよう配慮されている。通常の印字処理は、ほぼ、カバーできると思われるが、他に約 1 K 字分を予備として国研独自で選択した文字も登録可能となっている。ここでハード上の制限として、磁気ドラム上に格納した 3 K 字について、同一ページ内で印字可能な文字種は最大 24 字種までである。これは、ドラムのアクセスが遅いためプリンタ内部にある印字用バッファとの間に中間のドラム用バッファが必要となり、この容量が現在 24 文字となっているためである。（バッファの増設により最高 32 文字まで拡張可能である）実際の処理では、ROM 内の使用しなかった文字と一部入れ変えによって、最終印字段階までには、対処可能である。その他、このゲタ扱いとなった文字に対しては、新規文字のデザインを行う。また、ゲタ表示された●印の上に手書文字をはりこむ等も対応策の一つとなる。図 1 にこれらのコード体系と各領域内の漢字セットの配当を示す。

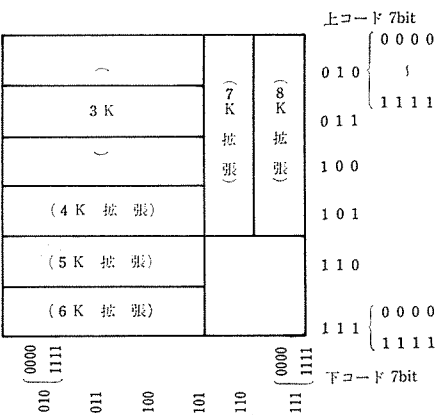
国研コード対 NEC コード変換テーブルについては、漢字パターン 3 K 字分と同様、磁気ドラム上に登録され、コードコンバートに使用される。ここで、印字の対象となる文字 8 K 字について、漢字テレタイプの盤内字、盤外字とも全てコードの対応付けが行われており、入力データのイメージどおりの印字（盤外字については、コードイメージのまま出力）また、盤外字変換イメージ、いずれの場合も処理可能である。これは、従来の漢字テレタイプのモニターでは、

図 1 キャクタコードセット

キャラクタ・セット



コード・セット



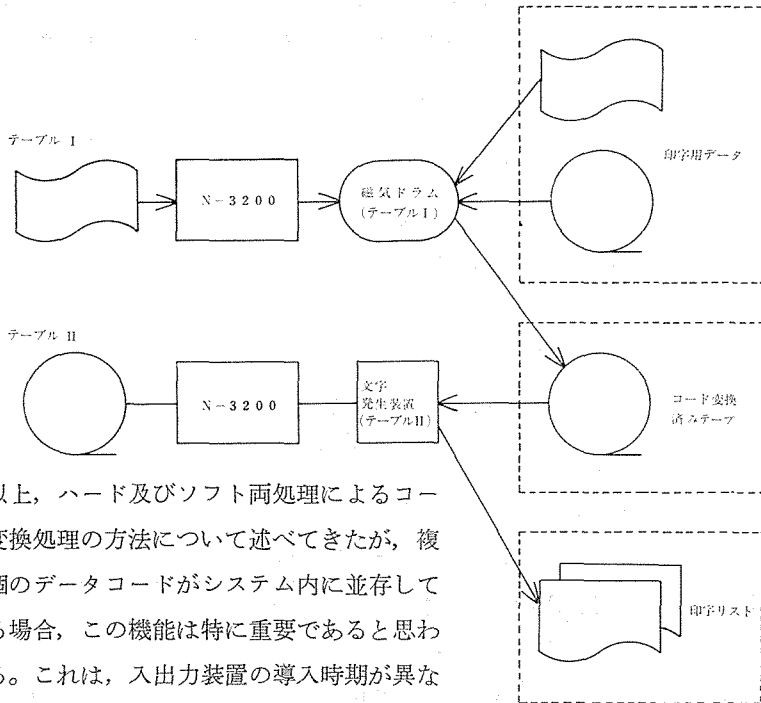
入出力は、一対一の対応であったため、全て入力イメージで出力される。これに対し、変換モードでの出力は、盤外字のコード化のチェック、また印字フォーマットの行ぞろえも簡単に処理できることになる。

コード変換テーブルの作成については、NEC コード16進4桁を漢字テレタイプの印字文字の後に付加し、1コードとする。盤外字の場合は、国研のコード化の方法、ダイヤモンドの後に盤内字ノンシフト400字の2文字の組み合わせを使用しコード化する。磁気ドラム上の格納形式は、ハッシュコードによる探索を行うため、国研コードによる下降順ソート後格納する。割りあては、外字探索時間の短縮を図るため、二領域に分割し、5632字のうち2048字をドラムの第一領域に、残る3584字を第二領域に分割配分している。探索方法はバイナリサーチを行っているが、単一領域の場合より効率はよい。

以上、テーブルの概要と使用装置について述べてきたが、入力レコードのビット構成を適当に使用することによって、半ハード的にコード変換処理を行うことができる。これは、国研コードの場合、6ビット2例で一漢字をコード化しているため、盤内字2400字分コード12ビットで可能な表現は4096となる。これをROM内の漢字パターン2400字分に対応させ、国研コード順に再配列後ROMに登録する。この場合、ROMの内容は国研コード順になっているため、直接国研コード順による印字が可能である。この場合、盤内字のコード変換処理は不用となる。現システムのROM領域が5K字分となっているのは、この理由による。漢字コードが12ビットまたは14ビットであれば、他のメーカーの場合も適用可能と思われる。（個々のコードの最小値が異なる場合、演算を必要とする場合もある。国研コードの場合16進（2020）を加える必要がある。この限りにおいては、完全なハード処理ではない）

これらの漢字辞書（テーブルⅡ）コード変換対応表（テーブルⅠ）を使用して、印字処理を行う過程と、各々のテーブル登録については、図2のようなデータ処理を必要とする。

図 2 漢字コード変換処理



以上、ハード及びソフト両処理によるコード変換処理の方法について述べてきたが、複数のデータコードがシステム内に並存している場合、この機能は特に重要であると思われる。これは、入出力装置の導入時期が異なった場合を含め、使用者のニーズにあった装置をアレンジメントした場合も、このコード体系の調整の問題が生じるためである。その他の漢字辞書の使用方法として、機種変更に伴う蓄積データの維持に、そのまま使用可能となる。辞書が確保されている限り、ある時間、ある装置のもとで作成されたデータに関しては、辞書の中にそれらの対応コードが残されているため、機種変更時点でのコード変換処理の必要はない。従来、漢字辞書の主な使用法は、異機種間のデータ変換のテーブルとして使用されてきたが、これは、データの保存と維持にも使用可能である。特に、国研の場合、データの蓄積は研究運用上不可欠であり、データ管理と運用の容易性は非常に重要である。

また、漢字辞書のメンテナンスについては、辞書中の付加情報を利用できるため、追加、削除等の項目の修正を通常の漢字処理の一部として処理でき、音

訓の読み、部首コード等の配列順のコードブックの作成も容易である。また、メーカー側から提供される管理プログラムの使用も可能となる。表2，3は漢字辞書の項目とコードブックの内容である。

表 2 漢字テーブル

図	番	シーケンス番号
漢 字 コ ー ド		NEC漢字コード (16進4桁)
ユ ー ザ コ ー ド		国研漢字 (8進4桁)
部	首	部首コード (10進3桁)
画	数	漢字の総画数
分	類	当用, 補正, 人名漢字等の判別記号
級		漢字セットの領域番号
漢 字 の 読 み		音訓別の代表的な読み
パ タ ー ン		漢字のドットパターン

表 3 漢字テーブルコード表

文 字	部首	画 数	分類	領 域	NEAC コード (14ビット)	NEAC コード (12ビット)	ユーザー コード	音読	訓 読	備 考
暇	072	13	T	3	2A2D	1215		カ	ヒマ	713279
過	162	12	T	3	2A2E	1216		カ	スギル	716875
千	051	03	T	3	2A2F	1217		カン	ホス	712400
刊	018	05	T	3	2A30	1220		カン	キザム	711217
汗	085	06	T	3	2A31	1221		カン	アセ	713802

5. ホストコンピュータの概要

現在のコンピュータ HITAC-8250 は HITAC-3010の後続機種として設置された。これは、当時の3010が記憶容量も小さく磁気テープを中心としたバッチ処理形態であったため、ジョブと利用者の増加とともに、これに対処しきれなくなったためである。コンピュータ規模が小さいことは、より高度の言語情報処理を目ざす場合、大きなネックとなる。少なくともアルゴリズムの研究開発から実用化のレベルへの拡張は不可能であった。このような背景から、新設

のコンピュータは、記憶装置の増設、特に周辺装置としてのランダムアクセス装置と記憶容量については最小限の要求事項として計画された。これによって、複数のジョブを同時に並行処理することも可能となり、従来単一ジョブで入出力、ソート等で専有されていた磁気テープは各領域単位に使用できるようになった。これは、ディスクによるバックアップがあつて可能となったものである。ディスクに対しては、種々の辞書類の格納とオペレーティングシステム領域として使用する。なお、領域1は40KB、第2領域は30KBである。

中央処理装置 HITAC-8250 (98KB)

磁気ディスク装置 (DISC) 29MB×2

磁気テープ装置 (MT) 40KB×6

カード読み取り装置 (CR) 1000枚/分

カードさん孔装置 (CP) 160桁/秒

紙テープ読み取り機 (PTR) 500字/秒

紙テープさん孔機 (PTP) 110字/秒

高速行印字装置 (カナ付) (LP) 1250行/分

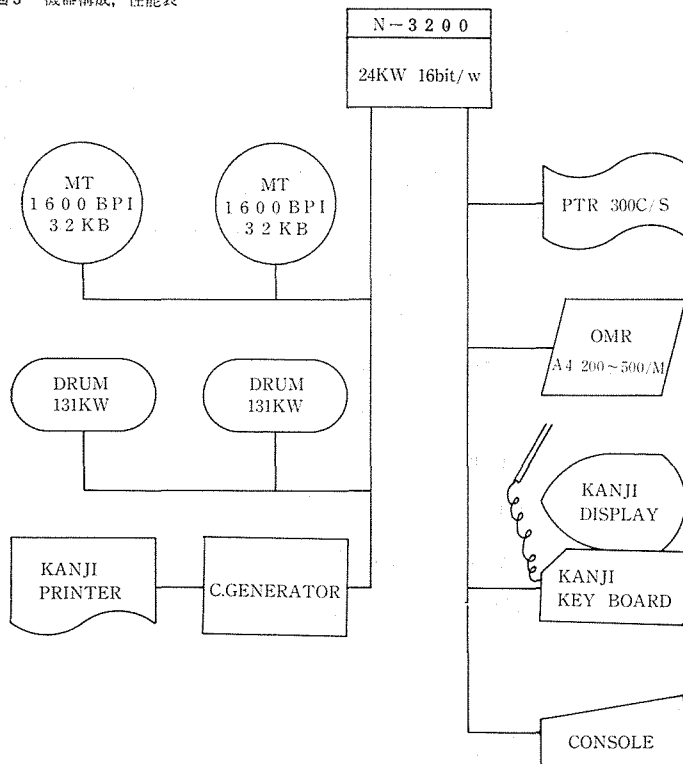
オフライン機器

カードけん盤せん孔装置 2台

紙テープ用タイプライタ 1台

上記の装置は、通常の汎用コンピュータ機器構成そのものであるが、ここで後述する、OMRカードによるターンアラウンドに必要なCRにはマーク読み取り装置は付加されていない。そのため、現状では、完全なターンアラウンドを構成することができないが、拡張機能として、付加できるため、現在では間接的、接続形式をとっている。その他、使用言語については、大部分は、COBOLによっているが、一部、統計業務についてFORTRAN, PL/Iを使用している。これらのコンパイラ言語は、言語処理にかならずしも最適であるとは言えない面が多い。特に、旧型機種 HITAC-3010 がキャラクタマシンであったため、基本的命令語の中に、バイトマシンにない機種がいくつかあり、それらが

図3 機器構成、性能表



取 容 文 字 (基本+外字)	4977+3584	記 録 方 式	多針電極 湿式静電記録
文字発生方式	ドット (24×24)	印 字 速 度	6 行/インチ 1800 行/分 3 行/インチ 1200 行/分
文字の大きさ	9, 12 P	用 紙 巾	200~457mm
文字記憶素子	ワイヤメモリ	用 紙 厚	55kg~135kg
印 字 行 数	80字×48行 (9, 12共)	用 紙 送 り	スプロケット
文 字 間 隔	9 P 7/インチ 12 P 5/インチ	用 紙	静電記録紙
行 間 隔	3/6 行インチ 4/8 行インチ	コ ピ ー 枚 数	同一ページ繰返印字
字 体	ゴチック, タイボス		

コンピュータに対する我々の母国語になっているため、この傾向が強い。その点、現在、種々のコンパイラが開発また提案されており、いわゆるコンパイラの世代の交代時期の様相が出始めている。しかし、汎用性を持った、機能拡大に従って、より大型の機械が必要になる場合が多い。例えば、PL/I 等についても、小型の場合、その制限が多くなり、従来の COBOL、FORTRAN との共通機能を中心となり、あまり PL/I の特長を示す命令は組み込まれていない。また、ファイル管理を目的とした、いわゆるデータベースソフトも弱体である場合が多い。

6. ソフトウェア・システムの概要

従来の用語用字調査は磁気テープによるバッチ処理方式であったため、入力データ作成に多くの経費と期間が必要であった。それは、調査対象となるデータの大部分が、コンピュータ入力以前に人手で処理、集計、編集、分類等の比較的単純な作業についてコンピュータを使用していたにすぎなかったためである。一つには、コンピュータそのものが通常の事務処理向きに開発されたものであり、周辺装置等も含め、漢字処理用に特に、開発されたものでないこともその大きな要因であった。また、当時、漢字を中心とした、言語情報処理は、まだ研究段階のものが多く、一部、新聞社等の活字鋳造機に接続させて、漢字テレタイプを使用していたにすぎない状態であった。その他、言語研究そのものも、機械に直接利用可能な形になっておらず、コンピュータ要員の確保とともに、システム開発を並行して進めなければならなかったこともその要因となっている。以上の点から、当初のシステムに多くを求めることは、無理であったことは否めない。しかし、最近、周辺装置の開発状態とミニコンピュータ等の比較的簡単に使用できる装置の普及、また、汎用コンピュータでの使用パターンの定着等から、漢字処理用端末装置の開発が進み、従来、人手作業にたよらざるをえなかった部分について、より緊密な共同作業が可能となった。本稿で述べるソフトウェアの基本と応用によって、従来から研究者がとってきた、

分析方法に近いやり方が適用できるようになったことは、これら作業に対処できる装置の開発におうところが多い。ここで、提案する応用システムは、入力装置 OMR、漢字プリンタ間の共通化した入出力媒体に情報を乗せるターンアラウンド処理である。図 15 はこのシステムの全体を示したものであり、図 5 は、入出力装置とプリンタ及びホストコンピュータのデータ経路の相互関係を示したものである。これらのシステムを従来のシステム構造と比較した場合、表 4 のような構造で表わすことができる。

Post-edit Type		Pre-edit Type	
人	パ ン チ 校 正	人	単 位 切 り 校 正
Comp	Computer	人	清 書
人	位 単 切 り 校 正	人	付 加 情 報 校 正
Comp	Computer	人	清 書
人	付 加 情 報 校 正	人	パ ン チ 校 正
Comp	Computer	Comp	Computer

表 4 システム構成

ここで、前処理型と後処理型の二つの型は、前者では、機械処理の前に、人手処理がまとめて行われており、特に、各付加情報作業の後にコンピュータ入力のためのデータさん孔が行われている。それ故、これを 前処理型 (Pre-edit Type) と名づける。また、後者の場合、個々の作業単位は、人手作業とコンピュータが対になっており、コンピュータ処理全体からの構造では (単位切り作業を含む)、付加情報処理がデータさん孔後に段階ごとに処理されており、前者に対し、これを 後処理型 (Post-edit Type) と名づけることができる。また、ここで述べる、ターンアラウンド処理は、この後処理型に属し形式的には、

$M \rightarrow mN$

$M \rightarrow m$

M, N : コンピュータ処理

m : 人間段階の処理 (研究も含む)

と表示できよう。これは、(人間処理←人間処理+コンピュータ処理)の反復構造となるため、コンピュータの前部分に相当する、人手作業を中心部をとりまく一つの環としてみるのが可能である。同様に、人間—機械系のインターフェースとしての入出力媒体もまた、これらの環構造(図4)と中心部を接続する環表現の一部となる。これらは、新たに追加される、新規の作業に対し、インターフェースとなる媒体形式がその使用目的に適合するかどうかの確認問題のみとなる。また、装置の増加の場合、ターンアラウンド機能を満足するためには、コンピュータとのインターフェース用媒体と他の作業の接続を示す、環上での作業帳表との条件を満足することが必要となる。例えば、80欄カードイメージで出力された用例カードの使用として、同語異語判別処理にそのまま使用できると思われるが、この場合、判別処理作業に適当な大きさの用紙であるか、記入できる情報量が適当であるか、また入出力装置の規格を満足しているか等のチェックが必要となる。

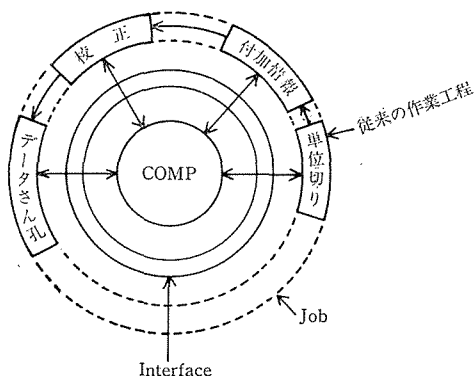
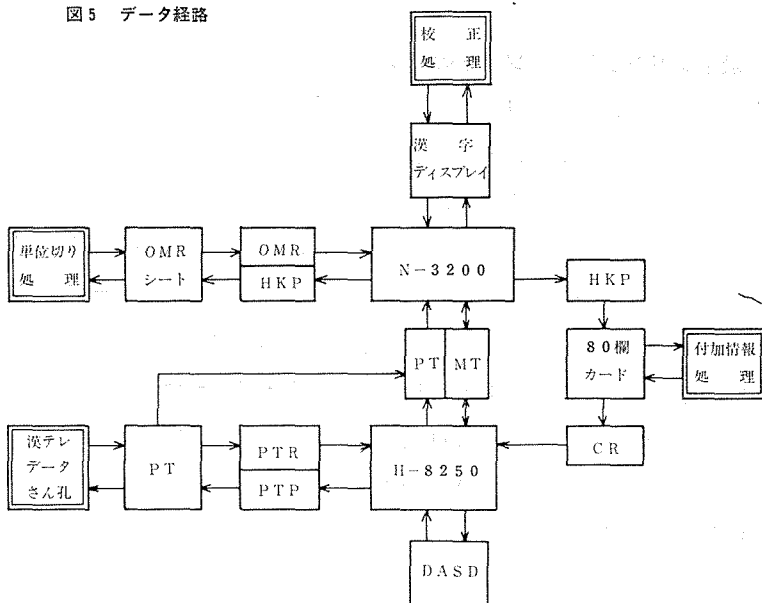


図 4 作業構造とジョブの流れ

これらの環上に配置された作業は、従来の作業の場合、直線的に作業環にそった形で進められ、データさん孔で、はじめて、コンピュータ処理に入る。これは、調査項目は付加情報作業としてコンピュータ入力以前に、全て処理されていなければならない。しかし、調査をデータ収集、データ処理、データ分析の各ブロックに分けた場合、分析後のデータ追加等の場合、データ収集ブロックから始めなければならないことを意味する。当然のことであるが、さん孔データのための原稿の作成、パンチ等、再調査と同様のシステム運営が必要となる。ターンアラウンド処理の場合、データ修正、追加データ等の原データへの接続は、カード、シートいずれの場合も事前の参照コードによって照合は容易である。また、作業手続きは全てカード上で処理できるため、調査システム全体は、作業者に対し、ブラック・ボックスは少なくなる。これは、入力媒体、出力媒体、作業帳表が同一媒体で統一されており、コンピュータとのインターフェースは作業者自身の目で確実に視覚化されているためである。また、各作業が階層的であることは、任意のレベルでのデータの使用が調査後についても使用でき、研究者は、調査と別に自身の方法で研究または、確認できることである。これは、原データを広く資料として、提供できるシステムと、各個人の研究を逆にシステムに接続できることになる。図5はホストコンピュータ、漢字プリンタのデータ経路を示したものであり、図4の環構造と各作業の対応を実際のコンピュータ上に展開したものである。付加情報挿入については、インターフェースに80欄カードを使用し、単位切り処理には、OMRシートを使用する。また、ホスト側とは、前述で示したように、電氣的に結合されていない。これらの利点は次のとおりである。

- 1) 第一次入力データを減少させることが可能。
- 2) システムのモジュール化が可能。
- 3) 外部データとの接続が容易。
- 4) 中間段階でのデータ再入力のミスマッチング防止可能。
- 5) データの再パンチ量を減少させることが可能。

図 5 データ経路



OMR 光学記号読み取り装置
 HKP 高速漢字プリンタ
 C R カードリーダー
 PTR 紙テープ読み取り装置
 PTP 紙テープさん孔装置
 DASD 磁気ディスク装置
 M T 磁気テープ装置

- 6) データの現場での作成が可能。
- 7) 同語異語判別処理が通常の付加情報処理と同一時点で処理可能。
- 8) データの転記作業が省略可能。
- 8) 用例カードによる情報検索用カードとして使用可能。

7. 漢字プリンタ用基本ソフトの概要

基本ソフトは、漢字プリンタ機器装置を稼働させる最小限のレベルで構成されている。プログラムは、導入以前にすでにメーカで開発済みのもの、新たに開発しなければならないものの、二種に分けられるが、新規開発分については、プリンタ導入決定時にメーカ側と協議し、できる限り初心者にも使いやすい方法をとった。特に、機械操作に伴う、コンソールのオペレートについて、コンピュータからのメッセージの略記号を選択応答することによって操作できるよう配慮した。以下、図10～図15のシステム・フローチャートに従い、ソフトの概説を行う。なお、N-3200型ミニコンピュータ上で処理される全てのプログラムは、絶対型ローダ、ABS-LDRによって制御される。

TBL-LD70 (図10参照)

紙テープで作成された、国研コード、NEC標準コード対応表を磁気ドラムに登録するプログラムである。紙テープベースとしたのは、初期のテーブル作成時に発生する誤りの修正を容易にするためである。また、校正チェックリストを得る上で、登録されたテーブルに対しテーブル自身をデータとして処理することによって、登録漏れ、コード付ミスの発見が非常に簡単になるためでもある。

次に、漢字テレタイプの盤内字と盤外字用コード変換テーブルの紙テープフォーマットを示す。

盤内字

XCCCC C/R XCCCC C/R~XCCCC C/R~E/B

- └─NEC標準漢字コード
- 0~9 A~Fの16進で指定
- └─漢字テレタイプ盤面上の文字

C/R (改行コード) は国研コード (5441) を使用する。

E/B (END BLOCK (7672)) は紙テープの終了を示す。

盤外字

◇XXCCCC C/R XXXCCCC C/R~◇XXCCCC C/R~E/B

- └─NEC標準コード
- └─国研用盤外字用コード

なお、このテーブルについては、国研コードで下降順にソートされていることが必要である。

CVT-PT70 (図12参照)

漢字テレタイプで作成された、紙テープをNECコードに変換し、磁気テープに出力するためのプログラムである。コード変換処理は、盤内字2400字の場合と、盤外字のコード変換処理の選択が可能である。盤内字モードの指定は、入力データイメージのまま出力印字する時に使用し、通常印字の場合、盤外字変換モードを指定する。行当りの印字文字数は最大80字であるが改行指定の区切りコードを使用し、印字体裁を整えることができる。しかし、区切りコードが二重に連続して使用されている場合、2番目以後は無視される。これらの区切りコードは、テーブル登録時に指定するが、現在表5の内容のコードが登録されている。

特殊コードの種類と機能

1) 読みとばしコード

オールマーク (ビット1~7まで全てON) の場合、1フレーム単位で読みとばす。紙テープの先頭及びブロック区切りコード以後のスプロケットは1フ

レーム単位で読みとばす。

2) ブロック区切りコードとして使用できるのは、次の7種である。

表5 ファンクションコード対応表-1

No.	国 研		NEAC	
	文 字	コード(8)	文 字	コード(16)
1	空 白	7777	一	5122
2	復 改	5441	二	5123
3	(I)	5453	三	5124
4	(II)	5413	四	5125
5	(III)	5454	五	5126
6	EI	7652	六	5127
7	Ed	7652	七	5128

3) 漢字コード以外の半角英数字，カナ文字については，半角指定コード，解除コードを使用することができる (CVT-PT70 においては，英数字半角モード処理機能はない)。また，NEC 標準コード表に対応する文字がない場合はゲタコードが出力される。

表6 ファンクションコード対応表-2

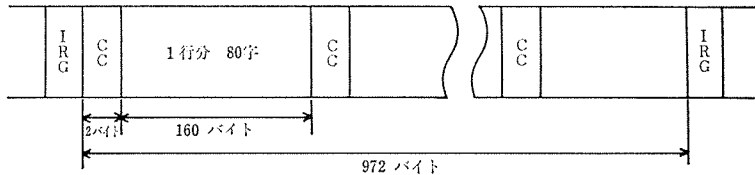
No.	国 研		NEAC	
	文 字	コード(8)	文 字	コード(16)
1	漢字IN(☆1)	1441	八	5129
2	漢字OUT(☆2)	1413	九	512A

ゲタ・コード

No.	国 研		NEAC	
			文 字	コード(16)
1	国研(漢テレ)to NEAC コード変換 テーブルドラムに登録されていない 文字		●	5027

CVT-MT70 (図12参照)

このプログラムの基本的機能は、CVT-PT70と同じである。HITAC 側で処理されたテープのコード変換に使用する。この出力テープは、MT-HKP-13, KDP-ED70, OMR-ED70の入力テープとなる。出力磁気テープ形式及びデータ長を次に示す。



6 レコード／ブロック

レコードサイズ 162 バイト

ブロックサイズ 972 バイト

1 レコードが80字に満たない場合は、(00)₁₆ でパディングする。

MT-HKP-13 (図11, 12参照)

CVT-PT70, CVT-MT70 で NEC 標準コードに変換された磁気テープデータを読み、漢字プリンタに出力するための汎用印字処理用プログラムである。ターンアラウンド用に編集されたデータについても、このプログラムを使用する。コマンドによって、入力データのレコード長、ページ当りの印字行数、コピー枚数、印字ラインの先頭位置等の指定が可能である。図16~18は印字出力サンプルである。図16は9 P, 6 行／インチ, 図17は12 Pで印字したものである。

KDP-ED70 (図11参照)

漢字ディスプレイを使用した校正処理のための前処理用プログラムである。この前処理は、主に、画面に表示するデータ単位のブロック化を行う。その他、修正済みデータから、修正分部のみを選択抽出する機能、表示のためにブロック化されたデータをデブロッキングする後処理用の機能も持つ。

漢字ディスプレイ C5310の性能

CRT（表示装置）12インチ 高解像度型

蛍光体 P39 緑色

表示文字容量 横書モード 26桁×16行=416文字

縦書モード 24桁×16行=384文字

英数字の場合、各々倍になる。

表示方式 漢字モード 17×18ドット

英数字モード 7×9ドット

画面 30回/秒リフレッシュ

KDP-CR70（図11参照）

漢字ディスプレイ及び漢字キーボードを使用し、KDP-ED70で処理されたデータ単位の検索、校正を行う。一処理ごとのデータ挿入、削除、置換は、画面表示されたコマンドに対し、ライトペンによる応答選択を行う。データブロック番号、英数字6桁の検索情報は漢字キーボードから指示する。（図6参照）

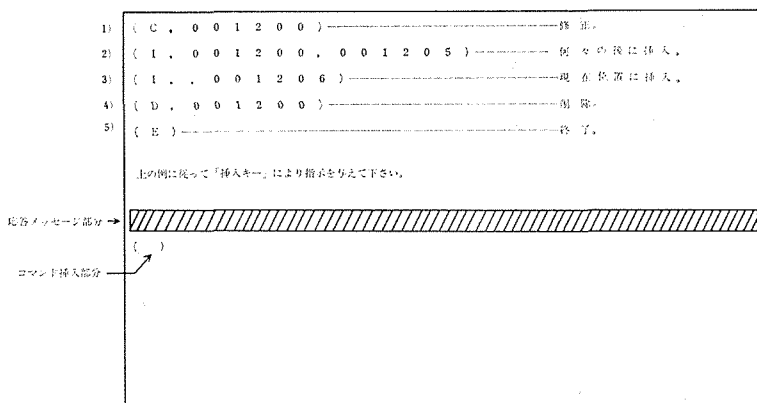


図6 コマンド入力要求メッセージ

1) (C, b_i) 修正コマンド

入力テープ上の各ブロックを b_i ブロックの直前まで出力テープにコピーし、b_i ブロックを画面に表示する。画面内容の修正終了指示を与えら

れると、 b_1 ブロックを修正済みデータとして出力テープに記録する。

2) (I, b_1 , b_2) 挿入コマンド

b_1 ブロックまでを出力テープにコピーし、 b_1 ブロックの後に新たに b_2 をブロック番号として登録し、キーボードからのデータを1ブロックとするデータの挿入を行う。

3) (I, b_2) 挿入コマンド

画面表示されているデータ上へ b_2 をブロック番号とする新たなデータを置き換える。ブロックの置換処理を行う。

4) (D, b_1) 削除コマンド

b_1 ブロックの直前までを出力テープにコピーし、指示された b_1 ブロックを削除する。

5) (E)

校正処理を終了させるコマンドで以下の処理は、入力テープの残りの情報を全て出力テープにコピーする。

OMR-ED70 (図13参照)

ターンアラウンド処理のため、NEC コード変換済みテープを OMR シートに合った印字形式に編集するためのプログラムである。実行時に、データ識別用コードを指定する。コードは、A～Eまでの重複しない英字の組み合わせと自動的にシート単位に付けられる3桁の数字である。このプログラムは、後述する、ターンアラウンド処理用応用プログラムの一部をなすものである。

MRK-ED70 (図13参照)

OMR-MT-20 で作成されたマーク情報を HITAC 側で処理可能な EBCDIC コードへ変換する編集処理と磁気テープラベルの調整を行う。マークシート上の情報との対応を図7に示す。

OMR-MT-20 (図13参照)

OMR シートを読み込みシート上のマークイメージを磁気テープに出力するプログラムである。基本的には、OMR シートを使ったターンアラウンド処理

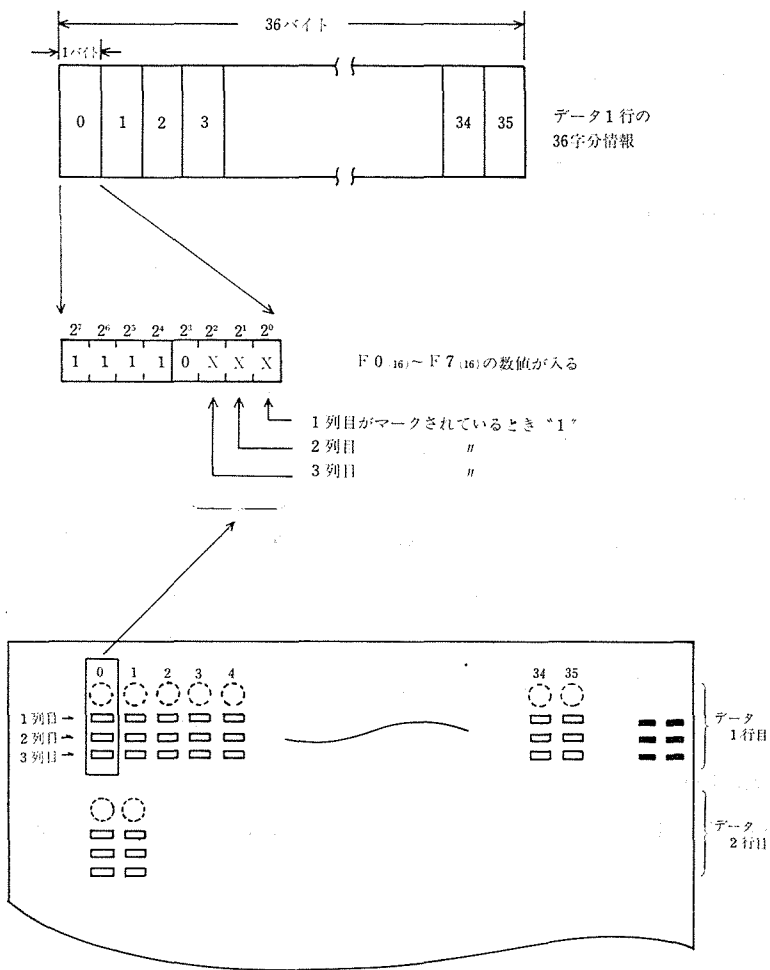
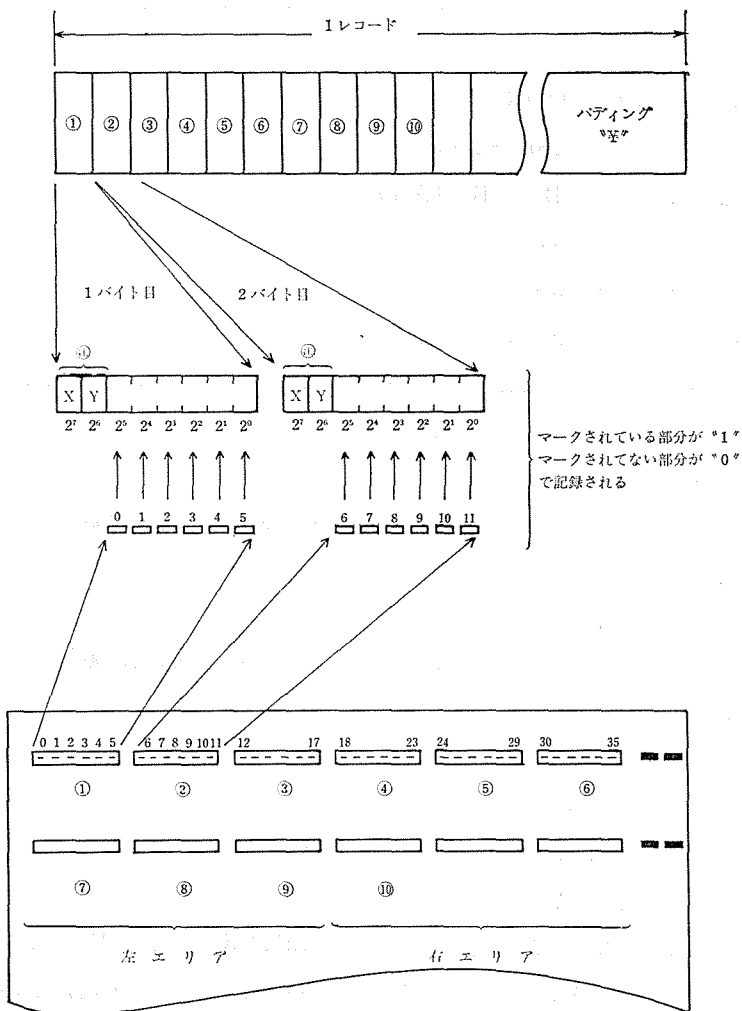


図7 OMRマーク位置と磁気テープ上の対応



① 読とりモードフラグが各バイトに付加される

X = 「1」のとき左エリアの情報「0」のとき右エリアの情報

Y = ビットモードで読まれたことを示すもので常に「0」である

図8 OMRマーク位置と磁気テープ1コードの対応

のために開発したものであるが、通常の OMR 処理にも利用できるよう、汎用性を持たせてある。

N240P-1 (B)	OMR 装置性能
読み取り速度	200枚～250枚／分
用紙サイズ	B 5 ～ B 4 (A 4)
マーク数	36個×66行
マーク・ピッチ	5.08mm
用紙規格	90～110kg用紙

なお、OMR シートの先頭及び終了を示す情報として、ヘッダー、トレイラー用シートを付けなければならない。

ヘッダ・シート

1 シートの最初のマーク30個のマーク状態が8進の“7700770077”であるシート。

トレイラ・シート

1 シートの最初のマーク30個のマーク数が8進の“0077007700”であるシート。

マークシートイメージの内容

2 レコード／ブロック

レコードサイズ 390バイト (1 シート分)

ブロックサイズ 780バイト

1 レコードが、390バイトに満たないとき、EBCDICコードの ¥(5 B)₁₆ がパディングされる。1 シートにおける最大情報量は、36 マーク 65 行=2340 マークである。1 レコードの記録内容とシート上のマーク位置の関係を図8に示す。

7. 応用プログラムの概要

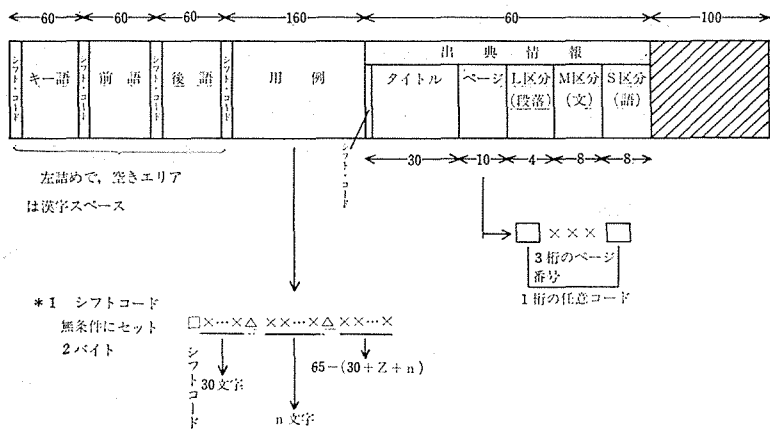
応用プログラムは、基本ソフトと対になって、ターンアラウンド処理を行う

ために開発されたもので、KKNMAT、KKNEDTの2本のプログラムから構成される。基本機能としては、プリンタ側機器をホストコンピュータで補うと共に、HITACで開発されたソフトウェアの利用を目的とする。KKNMATでの処理内容は、基本ソフトで処理された、OMR 付加情報テープと単位切りの対象となる原文データの照合マッチングを行い、単位切り情報の原文への挿入と KWIC イメージ磁気テープを作成する。また、KKNEDT では、KKNMAT で出力された KWIC イメージテープから印字イメージテープを作成するため、各項目に対し、削除、置換等の簡単な編集を行う。KKNMAT で出力された KWIC イメージテープは、100バイトの余白部分を予備エリアとして持っているため、データの追加、キー語、用例文を利用した、自動処理結果をフォーマット変換なしに付加することが可能である。その他、各プログラムは、できる限り汎用性を持たせるため、実行時にパラメータによる機能指定方式をとったが、一部、ブロッキングファクタ、レコード長等は、ソースカードのさし換えによる再コンパイルによって変更できるよう、機能の分散化を図った。これらは、プログラムリスト上でステートメントをアスタリスクで囲みその部分を明示してある。本システムで使用する装置は次の通りである。

	KKNMAT	KKNEDT
記憶装置	40KB	40KB (B: バイト)
磁気テープ	4 台	2 台
カードリーダー	1 台	1 台
ラインプリンタ	1 台	1 台

その他、パラメータカードの利用方法として、ジョブ単位で使用されるパラメータ内容と処理結果の履歴を磁気テープ等に記録しておくことによって、ジョブ管理またはデータ管理用情報として利用可能である。システム構成として図13～14にジョブの流れを示したが、HITAC 処理前の前部分は全て基本ソフトによってサポートされている。ここで応用プログラムとして述べるのは、図

14に示す点線で囲まれた部分についてである。また、80欄連続シートを利用した付加情報処理部分については、プログラム作成段階であるため、この中には含まれていない。図15はこれらを含む将来計画についてのシステムの流れを示したものである。次に、他の処理にも利用されることが多いと思われるKKNMATで出力されるKWICイメージテープフォーマットを図9に示す。



ファイル名	KWIC-IMAGE
レコードサイズ	500バイト
ブロッキングファクタ	2
パディング(斜線部分)	16進で00

図9 KWIC用例磁気テープフォーマット

7.1 KKNMATプログラムパラメータ機能

```

***  PARAMETER  CARD  ***                                     (51-03-26)
01_DSPEC2.....2031.....21342432.....2434.....2435
02_FCTCD.....42160.....2134.....512351275128

```

表7 KKNMATパラメータリスト

KKNMATで使用されるパラメータカードは、DSPCD、ECTD の二種である。以下、パラメータカードで使用される各項目について、実際の使用例にそって述べる。

DSPCD カードの使用例

表 8 DSPCDカードパラメータ

項 目	カラム数	説 明	使用例
シーケンスNo.	1・2	パラメータカードのシーケンスを指定	“01”であること
I D	4～8	パラメータの I D を指定	“DSPCD” であること
単位切 表示用 漢字 コード	単一条件	11～14 上欄 (“4”) にマーキングされている語のうしろに、ここで指定される漢字コードが挿入される	“2631” 漢字では○
		15～18 中欄 (“2”) にマーキングされている語のうしろに、ここで指定される漢字コードが挿入される	スペース
		19～22 下欄 (“1”) にマーキングされている語のうしろに、ここで指定される漢字コードが挿入される	“2632” 漢字では◎
	複合条件	23～26 上・中・下欄 (“7”) にマーキングされている語のうしろにここで指定される漢字コードが挿入される。単一条件がある場合はそのうしろにおかれる	スペース
		26～29 上・中欄 (“6”) 以下同上	“2334” 漢字では☒
		30～33 上・下欄 (“5”) 以下同上	“2432” 漢字では #
		34～37 上欄 (“4”) 以下同上	スペース
		38～41 中・下欄 (“3”) 以下同上	“2434” 漢字では %
		42～45 中欄 (“2”) 以下同上	スペース
		46～49 下欄 (“1”) 以下同上	“2435” 漢字では◎

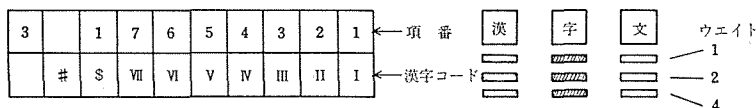
パラメータカードを使用する上で、シーケンス番号、識別用 ID コードについては問題ないと思われるが、ここでは、主に単位切り表示用コードの指定と OMR シート上の各マーク位置の関係について概説する。

(11—49カラム)

OMR シート上の各マーク位置は 3 ビット情報として個々に重み付けがなされているが、単独に扱う場合、また 3 ビットの複合条件として扱う場合の二条

件に分けられる。これらの条件選択は、パラメータカード上に指定される記号列とその位置によって決定されるが、条件の適当な組み合わせによって、単複両条件の併用指定も可能である。これは、後述するFCTCR カードのKWIC用キー語指定項目と直接関連するが、ここでは、FCTCR カード上で使用されている4, 2, 1, の各ウェイト設定条件を前提として説明を進める。

DSPCD 単位切り表示用条件コード



単一条件指定の場合

漢 字 \$#@

単複条件同時指定の場合

漢 字 \$#@ VII

複合条件記号

単一条件記号

単複条件同時指定の場合は、単一条件記号の後に複合条件記号が付加される。“VII”が選択されているのは、各マーク位置に付けられたウェイトの和が7であるため、7番目の記号が指定される。ここで、記号として付加される個数は単一条件3個、複合条件から1個、計4個である。

記号指定の必要のない場合は、スペースとし、各記号は NEC 漢字コード16進で指定する。また、単一条件を全てスペースとした場合、複合条件記号の“VII”が1個表示されるのみである。

以上の点から、実行パラメータ上では、FCTCD カードのキー項目がマーク位置の上位から1, 2, 4と指定されているため(13カラム1, 以下11, 12カラムは4, 2)1と2にマーキングされた場合

漢 字 ◎ %

となる。これは、単一条件の二項目がスペースであるためウエイト3の複合条件記号%が◎記号の直後に挿入されることになる。

FCTCD カードの使用例

表9 FCTCD カードパラメータ

項 名	カラム数	説 明	使用 例
シーケンスNO	1・2	パラメータカードのシーケンスを指定	“02”であること
I D	4～8	パラメータのI Dを指定	“FCTCD”であること
キー	11～13	上・中・下欄のどのマーキングをキー語として、扱うかの指定を行う（13カラム目はかならず指定する。4, 2, 1のいずれかであること）	“421”
文 字 数	14～15	H K P テープ作成するときの1行中の文字数を指定（“80”以下であること）	“80”
K用 W I C	タイトル	21～24 タイトルをはさむファンクション・コードを指定	“2334”
	ペ ー ジ	25～28 ページをはさむファンクション・コードを指定	
リス ト 用	改ページ	37～40 H K P テープ作成上、任意のところで改ページをさせたい場合、原文中あらかじめファンクション・コードを挿入し、ここで指定する	
	改 行	41～44 H K P テープ作成上、任意のところで改行をさせたい場合、原文中あらかじめ、ファンクション・コードを挿入し、ここで指定する	
キャリッジ・コントロール	61～80	原文中に任意のコードがあらわれたあと、スペースの場合このスペースを処理対象外とするためここで任意のコードを指定	“5123” “5127” “5128”

FCTCDカードでは、KWIC テープ作成のため、入力データ中の題名、ページの検出、OMRシートマーク位置のウエイトの設定、また、原文マーク挿入テープに対する印字形式を考慮した強制的改行、改ページ処理用ファンクション・コードの指定を行う。その他 OMRシート上の余白部分をデータ対象から削除するため、ブロック区切りコードと併用でファンクション・コード指定を行う。

(11～13カラム)

OMRシートの各マーク位置に対し、ウエイトの指定とキーを設定するための項目である。

11カラムからの3カラムに対しては、1, 2, 4の任意の数字の組み合わせが指定でき、13カラム目に指定された数がキーとなる。またキー項目 4, 2, 1, は文字の長さの大→小に対応しているため、3種の単位の中で最小単位がキーとして指定される。4, 1, 2として登録された場合は、キー語は2番目の長さの単位である中単位が指定される。

(14—15カラム)

漢字プリンタで印字する一行当りの文字数を指定する。

(21—28カラム)

KWICで使用される出典情報のうち、題名、ページ情報をデータから識別するファンクション・コードを指定する。ここで指定されたコードで囲まれた情報は、KWIC用出力テープ出典情報として記録し、用例文中には表われない。ただし、ページの場合、前後のコードを含め、5文字固定長、また題名の場合17文字である。

題名 / 星の王子さま /

ページ #007#

名々/(スラント=(2334)₁₆), # (イゲタ=(2432)₁₆)コードは、事前にFCTCDカード21~28カラムに NEC16 進コードで指定されていなければならない。

(37—44カラム)

原文マーク挿入プリント用強制改行、改ページコードの機能は、印字形式を整えるため使用する。

(61—80カラム)

のようなのでした。☺

□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□

OMRシート上に印字された文章が32文字以下の場合、文末以後に余白部分が生じる。これをデータ処理の対象外とするため、余白部分省略部にコードを指定する。指定可能なコードは、ブロック区切りコードとしてコード変換用テーブルに登録されている7種の中から任意に3個指定できる。コードが原文から検出されるとコードから次の文頭までのスペースが削除される。また、入力される原文テープ、付加情報テープのマッチング処理において、英字及び3桁の数字によるデータ識別コードチェックが行われるが、照合不可の場合次のエラーメッセージが出力される。

1) 原文テープ < 付加情報テープ

照合不可のデータを読みとばし、原文テープの識別コードをラインプリンタに印字する。エラーメッセージは下記の項目である。

シーケンス番号, UM-G, データ識別コード, 3桁の識別番号

2) 原文テープ > 付加情報テープ

ラインプリンタにエラーメッセージを印字する。エラーサインはUM-F。

3) 付加情報テープにダブリコードがある場合。

DUBL。以下同様。

4) 付加情報テープのキー語の長さが、フィールドの長さをオーバーした。エ

ラーサインはOV-F。以下同様。

その他プログラム終了時点で、入力データをカウントしリストを出力する。

表10 KKNMATデータカウントリスト

		レコード件数
入力レコード	COUNT (GENBUN)	324
	COUNT (FUKA)	14
出力レコード	COUNT (MARK)	18
	COUNT (KWIC)	257

7.2 KKNEDTプログラムパラメータ機能

表11 KKNEDT パラメータリスト

```

01 SPEC      80  20
02 HEAD      00KNJC
03           7C73  3F4F  3C33  3F25  3733  3427  2D37
04           264D2637264E
05
06           PG=1  5D5AEACH
07 HEAD      02
08           2948  2A48  3E36  32+F  434D  2221  3E4B  2521
09           434D  2851  2736  3536  2350  394C  3846  2345
10           2350  3536  2650EACH
11 HEAD      02
12
13           284D22362237213F21352238  113D2239213C2140223A  3C334226
14           324C434637223E4B394C422724FF04B7536  223A365A25532F3429332945
15           2C252938284EENDH
16 HEAD      0340482C29  3253  3F25
17
18           3D52
19           4151EACH
20 FORM 0103810102K 03850404K 239394K 1111665K
21 END

```

表12 SPEC カードパラメータ

項 目 名	カラム数	説 明	使用例
SEQ-NO	1～2	必ず数字で昇順でなければならないがSPACE の場合 CHECK されない。	0 1
ID	4～8	“SPEC△” と5桁で固定。	SPEC△
プリント文字数	11～12	1ラインにプリントする文字数で最大80文字までである。	80
プリント件数	15～16	1ページにデータを何行プリントするかを指定する99行まで可能であるがHKPにより44行までとなる。	20
改 ト ロ コ ン ル	左 端 位 置	強制的に改ページを行わせるための入力MT上のコードのカラム左端位置を10進数で指定。	指定して いない
	コード(16進)	強制的に改ページを行わせるための入力MT上のコードを16進で指定する。	指定して いない
イン プ ット レ コ ード サ イ ズ	51～54	COBOLプログラム内のKR-XのTIMESを指定したレコードサイズを確認するために4桁で指定し、人間の眼によりCHECK する。	0050

KKNEDT で使用されるパラメータは、入力データとなるKWICレコードを読み込み、項目の削除、変更によって、直接印字可能な印字イメージの磁気

テープファイルを作成することを目的とする。パラメータは四種あり、それぞれ印字形式の枠組指定用パラメータ、ヘッダー印字用パラメータ、入力データに対する項目削除及び印字項目への位置付け用パラメータ、また、カードの終了を示す END パラメータ等である。以下、パラメータの代表的な項目を説明し、実際に実行されたパラメータリストについて解説する。

SPEC カードの機能は KKNMAT の DSPCD カードと同様、出力印字イメージの大体の枠組みを指定する。

(15—16カラム)

1 ページに印字する行数を指定する。2 カラムで01から99まで指定可能であるが、

NEAC-C5210D 漢字プリンタのハード上の制限により、実際使用可能な行数は44行までである。また、ここで指定される行数の中には、ヘッダーで指定される行数は含まれない。

$$\text{ヘッダー部印字行数} + \text{用例文印字行数} \leq 44$$

の条件を満たすことが必要である。

(31—38カラム)

このカラムで指定されるコードは、KKNMAT、FCTCD カードで指定された強制改行、改ページと対になって使用され、原文単位切り記号挿入テープの印字調整用として用意されたものである。通常の KWIC 用例文を目的とする印字の場合、使用しない。

(51—54カラム)

入力レコード長を変更する場合、カードのさし換えと再コンパイルによって、任意の長さに変更することができるが、現在使用しようとしているレコードの長さをこのカラムで示す。プログラムによるチェックの対象とはなっていないが実行時の確認の意味を持つ。

パラメータ使用例では、最大印字行数80字、1 ページ当りの印字行数は20行とした場合の例である。

HEAD カードの使用例

表13 HEAD カードパラメータ

項 目 名	カラム数	説 明	使用例
SEQ-NO	1～2	上記に同じ	02～19
ID	4～8	“HEAD△”と5桁で固定。 複数語の指定が可能で2枚目以降はSPACEとする。	HEAD△
プリント ライン	11～12	00～15の改行数 (AFTER) を指定。(但し00の時は改ページを表す) HEADカードの一枚目に必要、2枚目以降はSPACEとする。	02 0R 03
H E A D I N G 用 漢 字 コ ー ド	13～16 17～20 21～24 25～28 29～32 33～36 37～40 41～44 45～49 } 77～80	通常、16進でNEAC漢字コードを指定するが次のコードについては特別の意味を持つ。 “△△△△”→2F20 (漢字スペース) “KNJC”→1B3E (漢字インコード) “ANKC”→1B3F (漢字アウトコード) “ENDH”→1ラインの最後に指定し、 後は漢字コードスペースとする。 “PG-n”→ページカウントを表示する (nは桁位置を示し1～5である)	

HEAD カードでは、プログラム実行時に任意にヘッダーを指定することを目的とする。指定可能範囲は、ページの先頭行から10行である。またプログラム内部でのページカウンタによって、5桁までのページ情報を、このヘッダー領域内に指定することができる。

(11—12カラム)

ヘッダー印字用の改行、改ページ情報を指定する。ただし、00の場合は改ページ扱いとなる。改行数の指定1～15は AFTER 処理となり、印字する前に指定行数をスキップする。

(13—80カラム)

カード1枚で指定できる字数は17字であるが、この中で、スペース、英数字モード、1ラインの終了コードページカウント表示については、ニーモニックな

記号を使用することが可能である。【使用例では、HEAD カードは、カード番号 2 ～19までの18枚であるが、2 枚目のカードで、このヘッダーが漢字モードで処理されることを示すKNJCが宣言されている。このヘッダーは表11のように印字されるが、内容は、プログラム作成を外部に委託したさいの納入資料の一部を使用したものである。16～19 枚目のカードで、KWIC 索引用の各項目、段階、語、用例の名称を指定している。また、5、6 枚目のカードで、PG-3、PG-2、PG-1はこの用例集のページ指定の送り項目を指定しており 5 FORM カードの使用例

表14 FORMカードパラメータ

項目名		カラム数	説 明	使用例
SEQ-NO		1 ～ 2	前記に同じ	20
I D		4 ～ 8	※FORM△と 5 桁で固定 3 枚まで可能で 2 枚目以降は SPACE とする。	FORM△
改 行 数		9 ～10	01～15の改行数 (AFTER) を指定。 FORM カードの 1 枚目に必要で 2 枚目以降は SPACE とする。	01
プリント情報	インプット 左端位置	11～14	入力データの編集する項目の左端位置	0381
	H K P 左端位置	15～16	編集された出力データのアウトプット	01
	プリント 文字数	17～18	編集する項目のプリント文字数 0 1 ～ (SPEC カードで指定したプリント文字数)	02
	プリント モード	19	そのフィールドの先頭におけるプリントモード “A”→ANKコード “K”または“△”→漢字コード	K
	区切り コード	20	この項目の区切りを示す	スペース
プリント 情報		21～29	上記のプリント情報に同じ	
		31～39	〃	
		41～49	〃	
		51～59	〃	
		61～69	〃	
		71～79	〃	

桁のページカウントから下3桁を指定したものである。これは、5桁のカウントから任意の桁が指定可能であることを示している。

FORM カードは、入力テープ上の各項目をプリントイメージに編集するための項目、移動桁数、送り先き、印字モードの4項目を各移動単位ごとに指定する。プリント情報は、カラム11～カラム20の10桁を1単位とし、カード1枚7項目、最大カード3枚21項目まで指定可能である。なお、出力印字イメージは、漢字プリンタの最左端より順次指定し、項目間をあけることはゆるされない。なお、カラム9、10の改行数は、印字される各行の改行指定であり、扱いは、1～15までの AFTER 処理となる。

(11～14カラム)

入力レコードの印字すべき項目の左端位置を指定。最左端位置は1である。

(15—16カラム)

プリンタ用紙の左端を1とし、指定された数字のカラムから入力データ項目が挿入される。

(17—18カラム)

入力データの転送項目からの転送文字数を指定。

(19カラム)

印字対象となる項目の印字モードを指定。

英数字モード：A 漢字モード：K

表10より。

03810102K	入力データの381バイトからの段落番号2文字分（英数字モードでは4バイト）をプリント領域のカラム1に転送。
03850404K	385バイトからの文番号項目4文字分をプリント領域のカラム4に転送。
03930904K	393バイトからの語番号項目4文字分をプリント領域のカラム9に転送。
01811665K	181カラムからの用例文項目65文字分をプリント領域のカラ

ム16に転送。

END カードの使用例

KKNEDT で使用されるパラメータの最後に END カードを挿入する。これは必須である。

表15 ENDカードパラメータ

項目名	カラム数	説明	現在の使用例
SEQ-NO	1～2	前記に同じ	21
ID	4～8	“END△△”と5桁で固定。 パラメーターカードの最後に必要。	END△△

8. 結び

漢字プリンタと OMR 装置を利用したターンアラウンドシステムの概要と有効性について述べてきたが、今後、漢字プリンタの開発についても従来の印字装置としての単能機能的装置から、情報または使用される媒体そのものについても再加工の道を考えるべきであろう。情報は、結局“乗り手”であり“乗せ手”である媒体についてもコンピュータ処理では無視できないものとなるからである。この点、システム的には一応成功したと思われるが、カード及びシートによる調査システムについて、問題となるのは、漢字の読みがな付け処理に効果的方法がないことである。これは、作表、語の配列処理等に直接影響を与えることになる。システムとしては、資料から得られる情報は可能な限り残し、それ以外の情報については、全て自動化の対象として扱っているためである。ここで提供されるデータは、第一次資料となる未加工データのみである。ただし、単位切り情報は、これら一次資料を使って得られる二次的資料であり、自動化処理またはカード等によって与えられる情報もこれに相当する。当然、読みがな処理は第二次以後の処理の中に含まれる。

その他、漢字プリンタ用ソフトウェアの開発に関しては、ハードウェアの導入後3段階に分け、最初に最も重要なプログラム、例えば、印字処理、紙テープ

コード変換用プログラムを開発し、第2段階目のプログラム納入までに、可能な限りチェックを行う方法をとった。これによって、第3段階のプログラム納入時には、重要プログラムほど、チェックされ、使いこまれる期間と回数が多くなりシステムの安定性は増すことになる。同時に、第3段階までに、各プログラムの不用部分の削除と必要な機能の挿入等プログラムの改良も可能となる。この方式の採用は、比較的新しいハードウェアのソフト面でのバックアップに対し通常の一括納入方式より効果的である。

最後に、このシステムの開発にあたって、ファコムハイタックの中島保行氏、大日本印刷の坂巻照夫氏、日本電気官庁システム部の方々に大変お世話になった、深く感謝の意を表する。

(この報告は、3月24日に岩波ホール集会室において、「漢字プリンタによるターンアラウンドシステム」と題して発表したものに手を入れたものである)

(1976.6.8.)

参考文献

- 1) 斎藤秀紀 漢字プリンタを使用したターンアラウンドシステム 国語研究所報告 51, 1—17 (1974)
- 2) 斎藤秀紀 漢字プリンタを使用したターンアラウンドシステムⅡ 国語研究所報告 54, 48—119 (1974)
- 3) 斎藤秀紀 国語研究所における高速漢字プリンタシステムの概要 ドクメンテーション研究 Vol 25, No. 8 (1975)
- 4) 特集 最近の電子映像記録法 電子科学 (1971)
- 5) 木練清蔵 ハードコピーと関連機器 工業調査会 (1973)
- 6) C-5210高速漢字プリンタ取扱い説明書 日本電気KK (1973)
- 7) NEACシリーズ2200周辺装置取扱い説明書 日本電気KK

図10 NEACコード変換テーブル・ロード

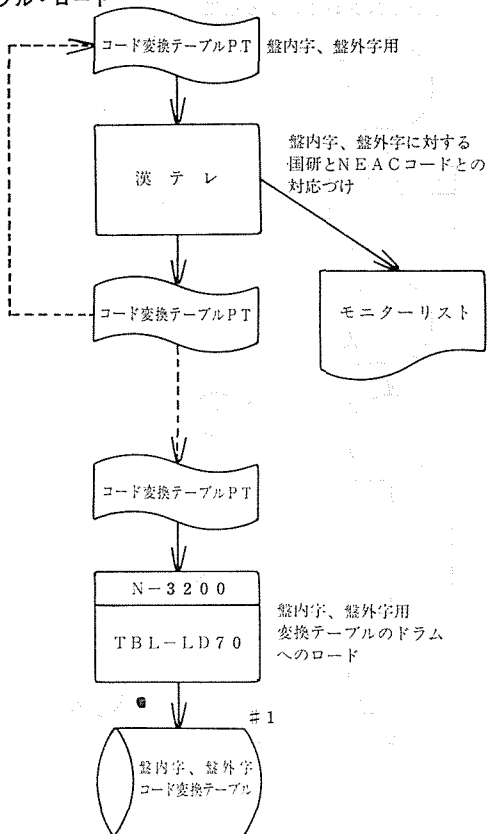


図12 コード変換および漢字プリント

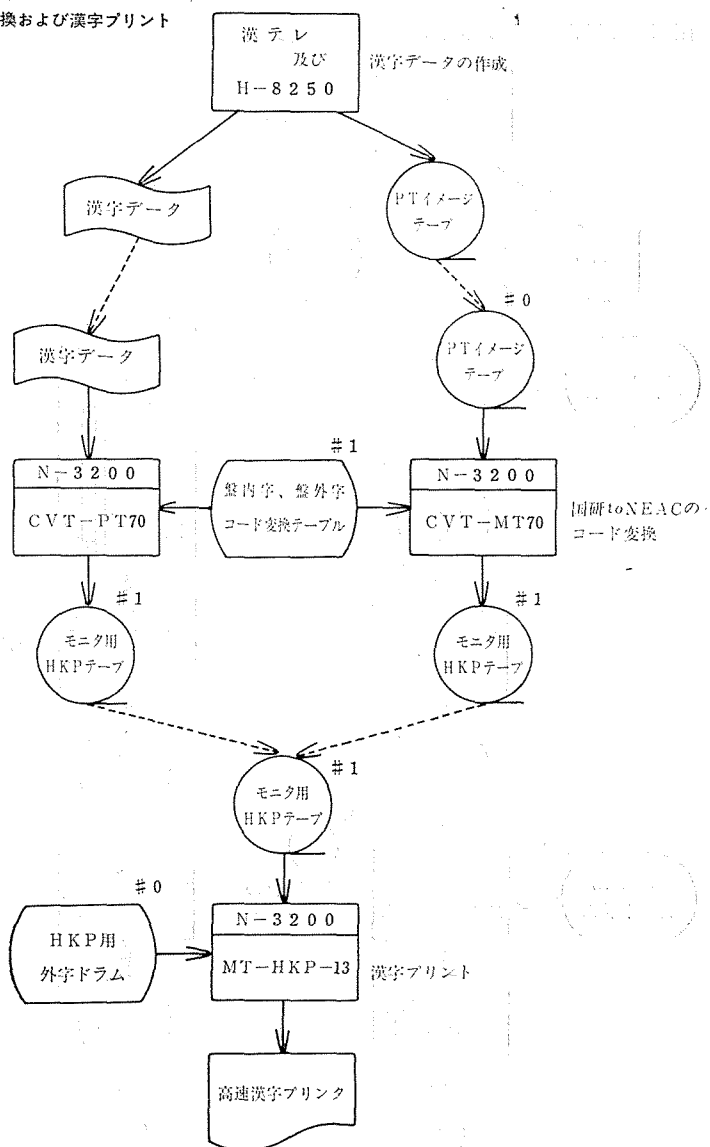


図13 OMRターンアラウンド処理の流れ

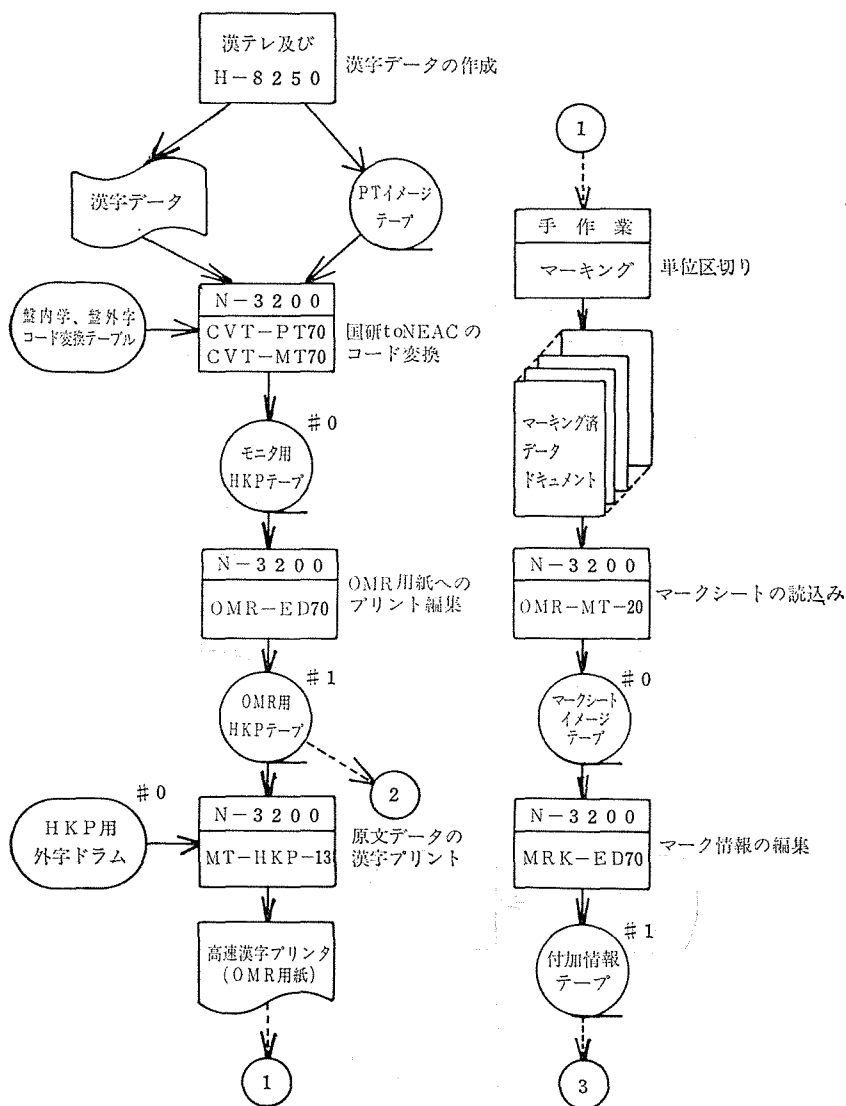


図14 OMR HITAC側処理

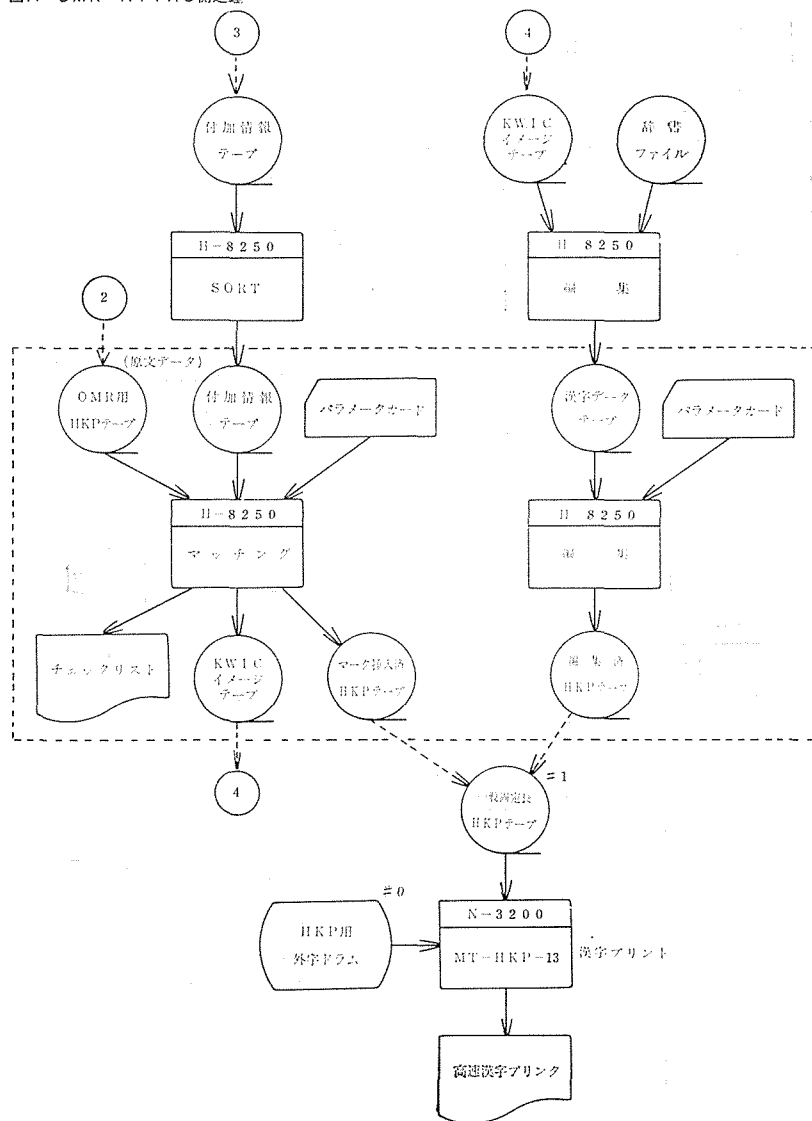


図15 ゼネラルフローチャート

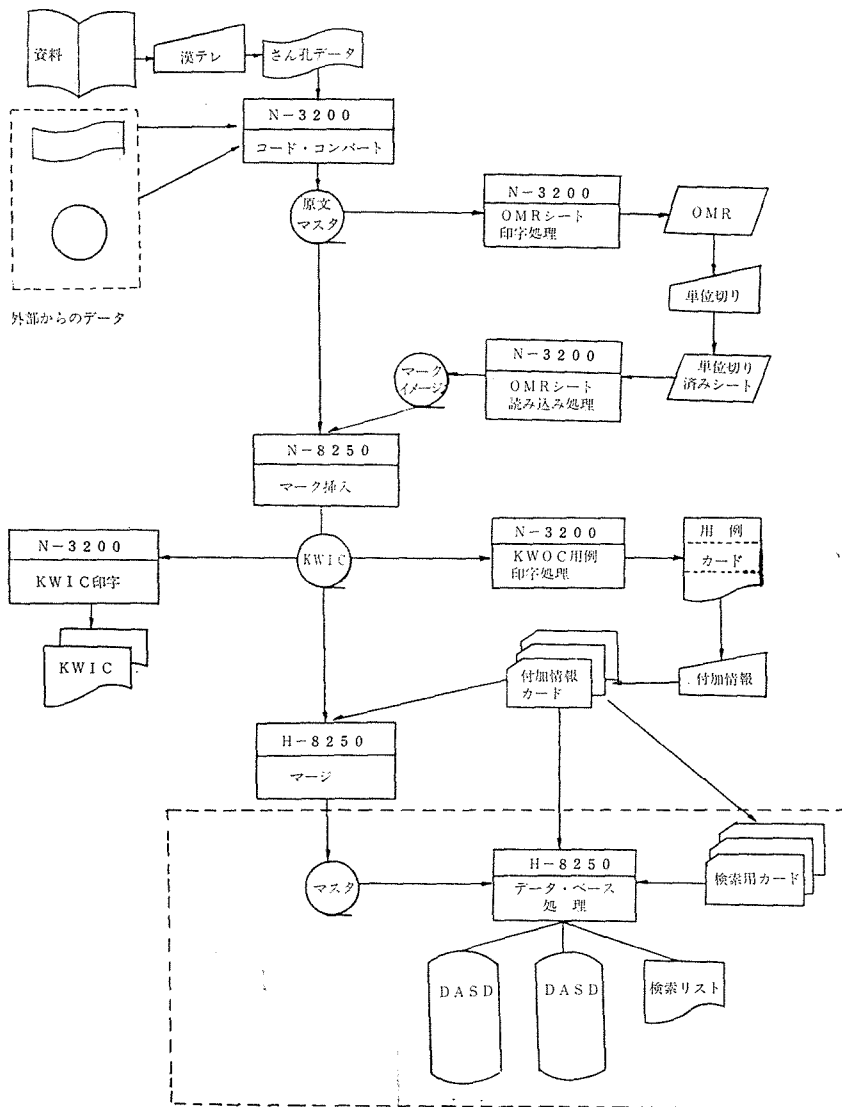


図16 漢字プリンタ印字例 (サンテグジュベリ作 内藤濯訳 星の王子さま 岩波少年文庫より)

／星の王子さま／

#7#

六つのとき、原始林のことを書いた「ほんとうにあった話」という、本の中で、すばらしい絵を見たことがあります。それは、一びきのけものを、のみこもうとしている、ウフバミの絵でした。これが、その絵のうつしです。

その本には、「ウフバミというものは、そのえじきをかまずに、まるごと、ペロリとのみこむ。すると、もう動けなくなって、半年のあいだ、ねむっているが、そのあいだに、のみこんだけものが、腹のなかでこなれるのである」と書いてありました。

ぼくは、それを読んで、ジャングルのなかでは、いったい、どんなことがおこるのだろうと、いろいろ考えて

#8#

みました。そして、そのあげく、こんどは、色エンピツで、ぼくのはじめての絵を、しゅびよくかきあげました。ぼくの絵の第一号です。それは、まえのページのようなのでした。

ぼくは、鼻をかかたと、その絵をおとなの人たちの見せて、〈これ、こわくない?〉とききました。

すると、おとなの人たちは〈ぼうしがなんてこわいものか〉といました。

ぼくのかいたのは、ぼうしではありません。ゾウをこなししているウフバミの絵でした。おとなの人たちに、そういわれて、こんどは、これなら、なるほどとわかってくれるだろう、と思って、ウフバミのなかみをかいてみました。おとなの人ってものは、よくわけを話さない、と、わからない。ぼくの第二号のようなのでした。

図17 OMRシート印字例

AB001 987654321098765432109876543210

／星の王子さま／

#7#

六つのとき、原始林のことを書いた「ほんとうにあった話」という、本の中で

、すばらしい絵を見たことがあります。それは、一びきのけものを、のみこもう

としている、ウフバミの絵でした。これが、その絵のうつしです。

その本には、「ウフバミというものは、そのえじきをかまずに、まるごと、ペ

ロリとのみこむ。すると、もう動けなくなって、半年のあいだ、ねむっているが

図18 80欄OMRカード印字例

植物	INDEX - NO.										情報I	情報II	情報III	情報IV	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
	4	4	4	4	4	4	4	4	4	4	4	4	4	4	
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
	6	6	6	6	6	6	6	6	6	6	6	6	6	6	
	7	7	7	7	7	7	7	7	7	7	7	7	7	7	
	8	8	8	8	8	8	8	8	8	8	8	8	8	8	
	9	9	9	9	9	9	9	9	9	9	9	9	9	9	
	付加情報処理カード														

さまざまの形をした***や動物が、基本的にはすべて同じ構造の細胞からなり、その本質的な化学成分もほぼ等しいことから、細胞が生物のからだを構成する基本単位ではないかという予想がうまれる

や	INDEX - NO.										情報I	情報II	情報III	情報IV	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
	4	4	4	4	4	4	4	4	4	4	4	4	4	4	
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
	6	6	6	6	6	6	6	6	6	6	6	6	6	6	
	7	7	7	7	7	7	7	7	7	7	7	7	7	7	
	8	8	8	8	8	8	8	8	8	8	8	8	8	8	
	9	9	9	9	9	9	9	9	9	9	9	9	9	9	
	付加情報処理カード														

さまざまの形をした植物***動物が、基本的にはすべて同じ構造の細胞からなり、その本質的な化学成分もほぼ等しいことから、細胞が生物のからだを構成する基本単位ではないかという予想がうまれる